View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Tilburg University Repository

Tilburg University

Semiparametric Duration Models

Drost, F.C.; Werker, B.J.M.

Publication date:
2001

Link to publication in Tilburg University Research Portal

Citation for published version (APA):
Drost, F. C., & Werker, B. J. M. (2001). *Semiparametric Duration Models*. (CentER Discussion Paper; Vol. 2001-11). Econometrics.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 12. May. 2021

CentER

Discussion paper

No. 2001-11

SEMIPARAMETRIC DURATION MODELS

By Feike C. Drost and Bas J.M. Werker

March 2001

ISSN 0924-7815

Semiparametric Duration Models

FEIKE C. DROST and BAS J.M. WERKER *
Tilburg University

March 5, 2001

Abstract

In this paper we consider semiparametric duration models and efficient estimation of the parameters in a non-i.i.d. environment. We show that, in the often-used Autoregressive Conditional Duration (ACD) model, the assumption of independent innovations is too restrictive to describe financial durations accurately. Therefore, we consider semiparametric extensions of the standard specification that allow for arbitrary kinds of dependencies between the innovations. The exact nonparametric specification of these dependencies determines the flexibility of the semiparametric model. We calculate semiparametric efficiency bounds for the ACD parameters, discuss the construction of efficient estimators, and study the efficiency loss of the exponential pseudo-likelihood procedure. This efficiency loss proves to be sizeable in applications. For durations observed on the Paris Bourse for the Alcatel stock in July and August 1996, the proposed semiparametric procedures clearly outperform pseudo-likelihood procedures. We analyze these efficiency gains using a simulation study which confirms that, at least at the Paris Bourse, dependencies among rescaled durations can be exploited.

KEYWORDS: Adaptiveness, Durations, One-step improvement, Semiparametric efficiency.

JEL-CODES: C14, C41

*CentER, Department of Econometrics, and Department of Finance, Tilburg University, P.O.Box 90153, 5000 LE, Tilburg, The Netherlands. We thank Joann Jasiak for kindly providing the data. Several helpful remarks have been given by Rob Engle and Neil Shephard at the $(EC)^2$ meeting in Madrid (1999).

1 Introduction

During the last decade, the availability of financial data at a tick-by-tick level has greatly increased. The irregularly spaced data requires new econometric techniques to extract the economic information contained in such data. This paper concentrates on the durations between transactions on financial markets. To that extent, we base ourselves on the seminal Autoregressive Conditional Duration (ACD) model of Engle and Russell (1998). For the data at hand, the traditional assumption of independently and identically distributed innovations seems to be inappropriate. Therefore, we consider semiparametric models imposing less structure on the innovations. To obtain efficient estimators in these semiparametric models, we have to extend the semiparametric results available from the emerging literature on semiparametrics.

During recent years, an enormous progress has been made in the area of semiparametric estimation. Starting with the work of Stein (1956), about the possibility of adaptiveness in the symmetric location model, the techniques have been further developed ever since. The work by Hájek and Le Cam is especially worth mentioning here. Traditionally, the models considered are based on i.i.d. observations. A fairly complete account on the state of the art in i.i.d. models can be found in the monograph by Bickel, Klaassen, Ritov, and Wellner (1993). Newey (1990) provides an overview from an econometric perspective. Semiparametric efficiency considerations and adaptiveness in time series have been discussed as well, beginning with Kreiss (1987a,1987b) for ARMA-type models. In this stream of literature the innovations are assumed to be i.i.d. Koul and Schick (1997) discusses nonlinear autoregressive location models with special emphasis on the initial value problem. Drost, Klaassen, and Werker (1997) considers so-called group models, covering nonlinear location-scale time series. Steigerwald (1992) studies linear regression models in a time series context. Linton (1993) discusses linear models with ARCH errors. Drost and Klaassen (1997) particularizes to the GARCH model and Wefelmeyer (1996) calculates efficiency bounds in models with general Markov type transitions.

The present paper drops the i.i.d. assumption on the innovations. The semiparametric techniques mentioned above are used and extended to build an adequate model for durations between transactions on financial markets. Therefore, we consider semiparametric specifications where the innovations may have dependencies of unknown functional form.[1] As shown in Section 2, such a specification leads to a nontrivial analysis of semiparametric efficiency. The empirical results in Section 4 show that the gain from considering the more complicated semiparametric procedures is sizeable. These efficiency gains are important since they allow for much more precise parameter estimates and predictions. This holds even in financial applications where the number of observations is typically large.

[1] We have chosen a model formulation which uses these innovations and do not specify the model in terms of hazard functions. Such a specification with hazard functions would be mathematically equivalent, but we found it easier to use the specification of Engle and Russell (1998).

The crucial ingredient in semiparametric efficiency calculations is the efficient score-function. Let us recall this concept here. For a rigorous treatment, one may consult, e.g., Bickel et al. (1993) or Drost, Klaassen, and Werker (1997). Consider a setup where i denotes the observation number and $\theta \in \Theta$ is a finite dimensional parameter of interest. We denote (conditional) expectations under θ by E_θ. In general, a score function $s_i(\cdot)$ is a random function of the parameter θ, such that

$$E_\theta \{s_i(\theta)\} = 0, \ \theta \in \Theta, \ i = 1, \ldots, n. \tag{1.1}$$

Generally, the expectation in (1.1) has to be conditional on "the past" in order to get a martingale structure allowing for the derivation of limiting distributional results of estimators based on s_i. (In a parametric setting the optimal score function is given by the derivative of the conditional log-likelihood for θ.) A Z-estimator $\hat{\theta}$ based on the score function s_i is defined as the solution of

$$\frac{1}{n} \sum_{i=1}^{n} s_i(\hat{\theta}) = 0.$$

Under sufficient regularity conditions, this estimator is asymptotically normal with influence function

$$D(\theta)^{-1} s_i(\theta), \tag{1.2}$$

where $D(\theta) = -E_\theta \dot{s}_i(\theta)$ and $\dot{s}$ is the derivative of s with respect to θ. Observe

$$D(\theta) = E_\theta s_i(\theta) \dot{l}_i(\theta), \tag{1.3}$$

where $\dot{l}_i(\theta)$ denotes the derivative of the conditional log-likelihood for θ. This property is well-known for the optimal parametric score $s = \dot{l}$ (in which case $D(\theta)$ is the Fisher information). An immediate consequence of the Hájek-Le Cam convolution theorem is that, in models that are regular in the sense that they satisfy the Local Asymptotic Normality property (LAN), optimal estimators have the influence function (1.2) based on $s = \dot{l}$.

The above analysis is, by construction, a parametric one. The key idea in a semiparametric analysis is to reduce the semiparametric problem to a specific well-chosen parametric one. This special parametric model is called the least-favorable parametric submodel (compare also Newey (1990)). For completeness we repeat the argument here. First, consider an arbitrary parametric submodel of the semiparametric model under consideration. Obviously, since the information for statistical inference decreases if one enlarges the model, lower bounds on the asymptotic behavior of estimators in the parametric submodel, are also lower bounds for the behavior of estimators in the complete semiparametric model. Hence, the supremum of the lower bounds over the class of all parametric submodels also gives a lower bound for the semiparametric model. The second problem is to prove that a given lower bound is sharp. Usually, sharpness of a given bound is proved by providing a semiparametric estimator attaining this bound. Hence, if one finds a parametric submodel and an estimator in the semiparametric model, such that the bound of the parametric submodel is attained by the semiparametric estimator, then the bound is sharp and the estimator is efficient. Consequently, the particular parametric submodel is a least-favorable parametric submodel.

In order to find the least-favorable submodel, a technique based on tangent spaces has proved to be very useful (see, e.g., Bickel et al. (1993) or Van der Vaart (1998)). If one passes from a parametric model (say a model in which the density f of the innovations is completely known), to a semiparametric model where one supposes that f is unknown, there is usually an efficiency loss. This efficiency loss is caused by local changes in the density f that cannot be distinguished from local changes in the parameter of interest θ. Let $\dot{l}$ denote the score function for θ in the parametric model. The tangent space for f is defined as the space generated by all possible score-functions that can be obtained by changes in the nonparametric nuisance parameter f. The least-favorable parametric submodel induces a nuisance score (i.e. an element of the tangent space) that is closest to the score $\dot{l}$ induced by θ. This nuisance element is, by construction, the projection of $\dot{l}$ onto the tangent space. The residual of this projection defines the information left for estimating θ once f is unknown. This residual is called the efficient score-function. In this paper we extend this idea to the situation where innovations are not likely to be i.i.d. (as in duration models). The known procedure for time series models with i.i.d. innovations is adapted to cover several forms of dependencies. In Section 2, we develop the necessary theory leading to the relevant tangent spaces and efficient score functions of the parameters of interest.

The paper's outline is as follows. In Section 2, we discuss duration models in their general form and we develop the semiparametric theory as discussed above for the non i.i.d. setting at hand. Examples (Section 2.3) show how common specifications may be obtained. These specifications include different assumptions on the innovations, like i.i.d.-ness or a Markov type assumption. The estimation problem is considered in Section 3. We consider the consistency and efficiency of pseudo-likelihood procedures and a construction generally leading to efficient semiparametric estimators. These semiparametric procedures prove to be superior over pseudo-likelihood procedures. Section 4 discusses the properties of the durations observed on the Paris Bourse for the Alcatel stock in July-August 1996. We choose this sample, because it has been considered previously in the literature, see, e.g., Gourieroux, Ghysels, and Jasiak (1999). To give a possible explanation for the semiparametric efficiency gains observed in Section 4, we study some parametric extensions of the basic ACD model in Section 5. These extensions are chosen such that they exhibit similar dependencies as we find in the Alcatel data. The simulation study in Section 5 confirms the empirical findings of Section 4. Finally, Section 6 contains some concluding remarks.

2 The ACD model

2.1 The parametric ACD model

In this paper we focus on the Autoregressive Conditional Duration (ACD) model as introduced in Engle and Russell (1998). Suppose that we observe durations $x_1, \ldots, x_n$. These x's represent the time elapsed between two events, e.g., transactions of some asset. Let $\mathcal{F}_i$ denote the information available for modeling $x_{i+1}, x_{i+2}, \ldots$. We will set $\mathcal{F}_i = \sigma(x_i, x_{i-1}, \ldots, x_0)$, but it is very well possible to include exogenous variables in $\mathcal{F}_i$.[2] The key ingredient in the ACD model is the (conditional) mean duration time,

$$E\{x_i|\mathcal{F}_{i-1}\} = \psi_{i-1}. \tag{2.1}$$

In its simplest form, the formulation of the ACD model is completed by stipulating, e.g.,

$$\begin{aligned} \mathbb{P}\{x_i \leq x|\mathcal{F}_{i-1}\} &= F(x/\psi_{i-1}), & (2.2)\\ \psi_i &= \alpha + \beta x_i + \gamma\psi_{i-1}, & (2.3) \end{aligned}$$

where F denotes a particular distribution function (or a parametric set of distribution functions) on the positive half-line. In this case, the parameter of prime interest is $\theta = (\alpha, \beta, \gamma)^T$. In its original parametric setting, standard choices of F include the exponential distribution and, as an extension, the Gamma and log-normal or Weibull distributions. The distribution F has to be normalized to have expectation one in order to identify the constant in the specification of ψ_i. If F is not specified parametrically, we obtain a semiparametric model. The model (2.2) is implicitly based on underlying i.i.d. innovations. It is not difficult to see that (2.2) is equivalent to saying that $\varepsilon_i = x_i/\psi_{i-1}$ defines a sequence of i.i.d. positive random variables, each with distribution function F.

The above model, including various extensions, is introduced by Engle and Russell (1998) and has been studied as well in Engle (2000) together with a modeling of prices. These papers explicitly recognize the fact that the independence assumption in (2.2) implies that all temporal dependence between durations is supposed to be captured by the conditional mean duration function ψ_i. In that case, several parametric and nonparametric specifications of the distribution of the innovations F are studied. In this paper, we relax the assumption of independent innovations. The general model is specified in the next section and analyzed subsequently.

2.2 The semiparametric ACD model

Often, the strong i.i.d. assumption (2.2) is considered to be unsuitable and one would like to relax it. In our specification, this is equivalent to allowing F to be dependent

[2] This is because the derivations that follow are independent of the parametric form of the conditional duration ψ_{i-1} defined in (2.1).

on the past as well. If it is unknown in what way F should depend on the past, a semiparametric approach seems to be the most reasonable one. We assume that one is willing to define a set of variables that may influence F and we will see that the actual choice of these variables influences the semiparametric analysis. In complete generality, we assume that $\mathbb{P}\left\{\varepsilon_i \leq \varepsilon | \mathcal{F}_{i-1}\right\}$ is $\mathcal{H}_{i-1}$-measurable, where $\mathcal{H}_{i-1} \subset \mathcal{F}_{i-1}$. So, the restricted information set $\mathcal{H}_{i-1}$ (of the full information set $\mathcal{F}_{i-1}$) defines the relevant past variables to be used as parameters in the conditional distribution of the innovations ε_i. As we will see, the situation where $\mathcal{H}_{i-1}$ is strictly smaller than $\mathcal{F}_{i-1}$ is both common and relevant. We do not assume that $(\mathcal{H}_i)$ forms a filtration, i.e., $\mathcal{H}_{i-1}$ is not necessarily included in $\mathcal{H}_i$.

Formally, our semiparametric model is now described by (2.1) and

$$\mathcal{L}\left(x_i/\psi_{i-1} | \mathcal{F}_{i-1}\right) = \mathcal{L}\left(x_i/\psi_{i-1} | \mathcal{H}_{i-1}\right), \quad a.s. \tag{2.4}$$

One may choose the specification (2.3) of ψ_i, but other choices (like the ones in Engle (2000)) do not change the arguments presented below. Writing $\varepsilon_i = x_i/\psi_{i-1}$, we clearly have from (2.1) that $E\{\varepsilon_i | \mathcal{F}_{i-1}\} = 1$. The choice of the restricted information set $\mathcal{H}_i$ formalizes the dependence among the innovations ε_i.

A model with independent innovations can be obtained by taking $\mathcal{H}_i$ equal to the trivial sigma-field. There are two other important cases. If one chooses $\mathcal{H}_i = \mathcal{F}_i$, one leaves the dependence structure of the ε_i completely unrestricted. In more familiar terms, this would lead to a model that is solely characterized by the moment condition (2.1). One could also set $\mathcal{H}_i = \sigma(\varepsilon_i)$. In that case, the conditional distribution of ε_i given the past, may only depend on ε_{i-1}. This induces a Markov assumption on the innovations. Of course, there are many more possibilities. The theoretical derivations in the rest of this paper are based on a general specification with an arbitrary choice of $\mathcal{H}_i$ and we will specialize to the above mentioned choices in order to point out their differences from an estimation point of view in Section 2.3.

In order to derive efficiency bounds in the semiparametric model described by (2.1) and (2.4) with an arbitrary specification of the conditional expected duration ψ_{i-1} and $\mathcal{H}_{i-1}$, we follow the steps as set out in the introduction. Let θ denote the Euclidean parameter of interest describing the functional form of the conditional mean duration ψ_{i-1}, for example $\theta = (\alpha, \beta, \gamma)^T$ in (2.3). Write f_{i-1} for the density associated with $\mathcal{L}\left(\varepsilon_i | \mathcal{H}_{i-1}\right)$. We assume that f_{i-1} admits a Radon-Nikodym derivative f_{i-1}'. Regularity conditions under which the results to be presented below hold, are standard in the semiparametric's literature (see, e.g., Bickel et al. (1993), Section 2.1, or Drost, Klaassen, and Werker (2000), Section 2).

The score function for θ can be obtained by differentiation of the log-likelihood:

$$\dot{l}_i(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta} \log\left(\frac{1}{\psi_{i-1}} f_{i-1}(x_i/\psi_{i-1})\right) = -\left(1 + \varepsilon_i \frac{f_{i-1}'(\varepsilon_i)}{f_{i-1}(\varepsilon_i)}\right) \frac{\mathrm{d}}{\mathrm{d}\theta} \log(\psi_{i-1}). \tag{2.5}$$

To obtain the efficient score function in the semiparametric model in which the conditional density f_{i-1} remains unspecified, we need to calculate the projection of the score $\dot{l}_i(\theta)$ on the tangent space generated by the nuisance function f_{i-1}. As is argued below along the general lines of, e.g., Bickel et al. (1993), this tangent space $T_i(\theta)$ is generated by all score functions $h_{i-1}(\cdot)$ for which:

$$\begin{aligned} h_{i-1}(\cdot) &\in \mathcal{H}_{i-1}, && (2.6)\\ 0 &= E\{h_{i-1}(\varepsilon_i)|\,\mathcal{H}_{i-1}\} = \int_\varepsilon h_{i-1}(\varepsilon)\mathrm{d}\mathbb{P}\{\varepsilon_i \le \varepsilon|\mathcal{H}_{i-1}\}, && (2.7)\\ 0 &= E\{\varepsilon_i h_{i-1}(\varepsilon_i)|\,\mathcal{H}_{i-1}\} = \int_\varepsilon \varepsilon h_{i-1}(\varepsilon)\mathrm{d}\mathbb{P}\{\varepsilon_i \le \varepsilon|\mathcal{H}_{i-1}\}. && (2.8) \end{aligned}$$

Condition (2.6) follows from the fact that f_{i-1} is known to depend on $\mathcal{H}_{i-1}$ only, so that scores obtained by local changes in f_{i-1} also depend on $\mathcal{H}_{i-1}$ only. Condition (2.7) is the standard constraint in tangent space calculations, following from the fact that densities, by definition, integrate to one. In more classical terms it represents the condition that expectations of score functions are always zero (compare (1.1)). Finally, condition (2.8) results from the moment restriction $E\{\varepsilon_i|\mathcal{F}_{i-1}\} = 1$. The argument is as follows. Local changes in f_{i-1} represented by the score h_{i-1} induce a change in the first (conditional) moment of $\int_\varepsilon \varepsilon h_{i-1}(\varepsilon)\mathrm{d}\mathbb{P}\{\varepsilon_i \le \varepsilon|\mathcal{H}_{i-1}\}$. However, this moment is restricted to be one by condition (2.1). Therefore, the change must always be zero (otherwise one would not remain in the specified model). With these ingredients, we are ready to state the key proposition providing the lower bound for estimation of the parameters in ψ_i of the general semiparametric model described by (2.1) and (2.4).

Proposition 2.1 *In the semiparametric model described by* (2.1) *and* (2.4)*, the projection of the score function* $\dot{l}_i(\theta)$ *in* (2.5) *on the tangent space* $T_i(\theta)$ *defined by* (2.6)-(2.8) *is given by* $h^*_{i-1}(\varepsilon_i)$ *with*

$$h^*_{i-1}(\varepsilon) = -\left(1 + \varepsilon\frac{f'_{i-1}(\varepsilon)}{f_{i-1}(\varepsilon)} + \frac{\varepsilon - 1}{var\{\varepsilon_i|\mathcal{H}_{i-1}\}}\right) E\left\{\frac{d}{d\theta}\log(\psi_{i-1})\middle|\,\mathcal{H}_{i-1}\right\}. \qquad (2.9)$$

Proof: First of all, note that the proposed projection (2.9) indeed belongs to the tangent space $T_i(\theta)$ since it satisfies conditions (2.6)–(2.8). Secondly, the residual of the proposed projection of $\dot{l}_i(\theta)$ can be written as

$$\begin{aligned} \dot{l}^*_i(\theta) &\equiv \dot{l}_i(\theta) - h^*_{i-1}(\varepsilon_i)\\ &= \frac{\varepsilon_i - 1}{var\{\varepsilon_i|\mathcal{H}_{i-1}\}} E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\middle|\,\mathcal{H}_{i-1}\right\}\\ &\quad - \left(1 + \varepsilon_i\frac{f'_{i-1}(\varepsilon_i)}{f_{i-1}(\varepsilon_i)}\right)\left[\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1}) - E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\middle|\,\mathcal{H}_{i-1}\right\}\right]. \qquad (2.10) \end{aligned}$$

We show that both terms at the right-hand side are orthogonal to the tangent space $T_i(\theta)$. Let $h_{i-1} \in T_i(\theta)$ be arbitrary. Then, we obtain for the first term:

$$E\left\{\frac{\varepsilon_i - 1}{var\{\varepsilon_i|\mathcal{H}_{i-1}\}} E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\middle|\,\mathcal{H}_{i-1}\right\} h_{i-1}(\varepsilon_i)\right\}$$

$$= E\left\{ E\left\{ (\varepsilon_i - 1) h_{i-1}(\varepsilon_i) \middle| \mathcal{H}_{i-1} \right\} \frac{E\left\{ \frac{\mathrm{d}}{\mathrm{d}\theta} \log(\psi_{i-1}) \middle| \mathcal{H}_{i-1} \right\}}{var\{\varepsilon_i | \mathcal{H}_{i-1}\}} \right\}.$$

From equations (2.7) and (2.8) we see that the latter term equals zero, proving the desired orthogonality.

For the second term in (2.10), we obtain

$$\begin{aligned}
&E\left\{ \left(1 + \varepsilon_i \frac{f'_{i-1}(\varepsilon_i)}{f_{i-1}(\varepsilon_i)}\right) \left[\frac{\mathrm{d}}{\mathrm{d}\theta} \log(\psi_{i-1}) - E\left\{ \frac{\mathrm{d}}{\mathrm{d}\theta} \log(\psi_{i-1}) \middle| \mathcal{H}_{i-1} \right\}\right] h_{i-1}(\varepsilon_i) \right\} \\
&= E\left\{ E\left\{ \left(1 + \varepsilon_i \frac{f'_{i-1}(\varepsilon_i)}{f_{i-1}(\varepsilon_i)}\right) \left[\frac{\mathrm{d}}{\mathrm{d}\theta} \log(\psi_{i-1}) - E\left\{ \frac{\mathrm{d}}{\mathrm{d}\theta} \log(\psi_{i-1}) \middle| \mathcal{H}_{i-1} \right\}\right] h_{i-1}(\varepsilon_i) \middle| \mathcal{F}_{i-1} \right\} \right\} \\
&= E\left\{ \left[\frac{\mathrm{d}}{\mathrm{d}\theta} \log(\psi_{i-1}) - E\left\{ \frac{\mathrm{d}}{\mathrm{d}\theta} \log(\psi_{i-1}) \middle| \mathcal{H}_{i-1} \right\}\right] \right. \\
&\qquad \left. \times E\left\{ \left(1 + \varepsilon_i \frac{f'_{i-1}(\varepsilon_i)}{f_{i-1}(\varepsilon_i)}\right) h_{i-1}(\varepsilon_i) \middle| \mathcal{F}_{i-1} \right\} \right\} \\
&= E\left\{ \left[\frac{\mathrm{d}}{\mathrm{d}\theta} \log(\psi_{i-1}) - E\left\{ \frac{\mathrm{d}}{\mathrm{d}\theta} \log(\psi_{i-1}) \middle| \mathcal{H}_{i-1} \right\}\right] \right. \\
&\qquad \left. \times E\left\{ \left(1 + \varepsilon_i \frac{f'_{i-1}(\varepsilon_i)}{f_{i-1}(\varepsilon_i)}\right) h_{i-1}(\varepsilon_i) \middle| \mathcal{H}_{i-1} \right\} \right\},
\end{aligned}$$

where the last equality follows from (2.4). It is easily seen that this expression equals zero. This completes the proof of the proposition. □

As mentioned before, the residual (2.10) of the projection (2.9) is the efficient score function, which we denote $\dot{l}_i^*(\theta)$. Optimal semiparametric estimators must be based on this score-function. However, (2.10) cannot be used directly, since it depends on the unknown density f_{i-1} and on $E\{(\mathrm{d}/\mathrm{d}\theta)\log(\psi_{i-1})|\mathcal{H}_{i-1}\}$. In Section 3.2, we discuss how to estimate f_{i-1} and $E\{(\mathrm{d}/\mathrm{d}\theta)\log(\psi_{i-1})|\mathcal{H}_{i-1}\}$ in order to get a semiparametrically efficient estimator of θ.

Adaptiveness occurs (by definition) in case the efficient score function (2.10) equals the parametric score function (2.5). Thus, adaptiveness means that the projection of the parametric score on the tangent space is zero. In that case, there is (asymptotically) as much information in the semiparametric model as in the parametric model for estimating θ: the parametric score and the semiparametrically efficient score coincide. In the ACD model (2.3), we have $\psi_{i-1} > 0$ which for that specification implies $(\mathrm{d}/\mathrm{d}\theta)\log(\psi_{i-1}) > 0$.[3] Hence, adaptiveness occurs if and only if

$$1 + \varepsilon_i \frac{f'_{i-1}(\varepsilon_i)}{f_{i-1}(\varepsilon_i)} + \frac{\varepsilon_i - 1}{var\{\varepsilon_i | \mathcal{H}_{i-1}\}} = 0.$$

[3]Use $(\mathrm{d}/\mathrm{d}\theta)\log(\psi_i) = \psi_i^{-1}(\mathrm{d}/\mathrm{d}\theta)\psi_i$ and $(\mathrm{d}/\mathrm{d}\theta)\psi_i = \beta(\mathrm{d}/\mathrm{d}\theta)\psi_{i-1} + (1, x_{i-1}, \psi_{i-1})^T$.

It is easily seen that this is equivalent to, for some $c > 0$,

$$f_{i-1}(\varepsilon) = \frac{c^{-1/c}}{\Gamma(1/c)} \varepsilon^{1/c-1} \exp(-\varepsilon/c), \quad \varepsilon > 0. \tag{2.11}$$

Hence, adaptiveness occurs if and only if the conditional innovation's distribution is of the Gamma type (rescaled to have expectation 1). A similar result has been obtained for location models where adaptiveness occurs for the normal distribution and symmetrized square-roots of χ^2-distributions (see Gonzalez-Riviera (1997)). In our scale case, we have adaptiveness for the exponential and Gamma-distributions. The practical consequence of such a result is, of course, limited since the bound is calculated in a model that does not make any distributional assumptions.

It is well-known that densities at which adaptiveness occurs, are often also the densities for which the pseudo maximum likelihood estimator (PMLE) is consistent (see, e.g., Bickel (1982)). This shows that, a PMLE type estimator is consistent if and only if it is based on a Gamma distribution. Since for these densities $1+\varepsilon f'(\varepsilon)/f(\varepsilon)$ is always proportional to $1-\varepsilon$, the obtained PMLE estimators are in fact identical and the resulting PMLE is purely based on the moment condition (2.1). The estimator thus obtained is consistent in the full semiparametric model. In Section 3.1, we will see that the PMLE is only semiparametrically efficient under very restrictive conditions. An alternative estimator that is semiparametrically efficient in the model under consideration is given in Section 3.2.

The information for estimating θ in the parametric model is given by the variance of the parametric score (2.5). Assuming stationarity, this yields

$$E\left\{ J_{f_{i-1}} \left(\frac{\mathrm{d}}{\mathrm{d}\theta} \log(\psi_{i-1}) \right) \left(\frac{\mathrm{d}}{\mathrm{d}\theta} \log(\psi_{i-1}) \right)^T \right\},$$

where J_f denotes the Fisher information for scale, i.e.,

$$J_f = \int \left(1 + \varepsilon \frac{f'(\varepsilon)}{f(\varepsilon)} \right)^2 f(\varepsilon) \mathrm{d}\varepsilon.$$

The information loss of the semiparametric model, with respect to the parametric model, is given by the variance of (2.9):

$$E\left\{ \left(J_{f_{i-1}} - \frac{1}{var\{\varepsilon_i|\mathcal{H}_{i-1}\}} \right) E\left\{ \frac{\mathrm{d}}{\mathrm{d}\theta} \log(\psi_{i-1}) \middle| \mathcal{H}_{i-1} \right\} E\left\{ \frac{\mathrm{d}}{\mathrm{d}\theta} \log(\psi_{i-1}) \middle| \mathcal{H}_{i-1} \right\}^T \right\}.$$

Note that the information loss is indeed zero (adaptiveness) if and only if the (conditional) density f_{i-1} belongs to the Gamma class. This follows, since we have, by the Cauchy-Schwarz inequality,

$$J_{f_{i-1}} \int (\varepsilon - 1)^2 f_{i-1}(\varepsilon) \mathrm{d}\varepsilon \geq \left[\int (\varepsilon - 1) \left(1 + \varepsilon \frac{f'_{i-1}(\varepsilon)}{f_{i-1}(\varepsilon)} \right) f_{i-1}(\varepsilon) \mathrm{d}\varepsilon \right]^2 = 1,$$

with equality if and only if f_{i-1} is of the form (2.11). The information in the semiparametric model is given by the variance of the residual of the projection which, by the Pythagorean theorem, equals

$$E\left\{ J_{f_{i-1}} \left(\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\right)\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\right)^T \right\}$$
$$-E\left\{ \left(J_{f_{i-1}} - \frac{1}{var\{\varepsilon_i|\mathcal{H}_{i-1}\}}\right) E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\middle|\mathcal{H}_{i-1}\right\} E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\middle|\mathcal{H}_{i-1}\right\}^T \right\}.$$

2.3 Examples

We consider the efficiency calculations in more detail in four specific models.

Example 2.1 [i.i.d. innovations] In case the restricted information set $\mathcal{H}_i$ is the trivial sigma-field, we obtain that $E\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1})\middle|\mathcal{H}_{i-1}\right\}$ is a vector of constants. This implies that all components of the projection (2.9) generate the same direction in the tangent space $T_i(\theta)$. Adaptiveness in such models is well studied, see Drost, Klaassen, and Werker (1997).

Example 2.2 [Markov innovations] In a true Markov setting of the innovations, one would take $\mathcal{H}_i = \sigma(\varepsilon_i)$. The efficient score (2.10) does not simplify in this Markov case. Therefore, general statements are difficult to make in this setting.

Example 2.3 An important simplification in the efficient score function is obtained if $\mathcal{H}_{i-1}$ contains $\mathrm{d}\log\psi_{i-1}/\mathrm{d}\theta$, i.e., $\mathcal{H}_{i-1} \supset \sigma(\mathrm{d}\log\psi_{i-1}/\mathrm{d}\theta)$. In that case, the second factor in (2.9) reduces to $\mathrm{d}\log\psi_{i-1}/\mathrm{d}\theta$ and the efficient score becomes

$$\frac{\varepsilon_i - 1}{var\{\varepsilon_i|\mathcal{H}_{i-1}\}}\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1}). \tag{2.12}$$

In this expression, the (conditional) density f_{i-1} enters only through $var\{\varepsilon_i|\mathcal{H}_{i-1}\}$. This shows that the semiparametrically efficient estimator of θ is the moment estimator based on (2.1) with (optimal) instrument, compare Wefelmeyer (1996),

$$\frac{1}{var\{\varepsilon_i|\mathcal{H}_{i-1}\}}\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1}).$$

Note that our general semiparametric approach shows that, in the present example, the optimal semiparametric estimator is a moment estimator. We did not limit attention to moment estimators a priori.

Example 2.4 [moment condition] Consider the case where $\mathcal{H}_i = \mathcal{F}_i$. Then, the efficient score function is as in the previous example, except that in this case $var\{\varepsilon_i|\mathcal{H}_{i-1}\} = var\{\varepsilon_i|\mathcal{F}_{i-1}\}$. Again, the optimal estimator is a moment estimator with above mentioned instruments and weighting matrix. The efficient score does not alter if one enlarges the model of Example 2.3 to a model in which no additional structure is imposed. One may also turn this argument around. Starting from a model which is solely characterized by the relation (2.1), no statistical information is added if one imposes that the conditional distribution of the innovations given the past $\mathcal{F}_{i-1}$ is determined by $\mathrm{d}\log\psi_{i-1}/\mathrm{d}\theta$ alone. In that sense, adaptiveness occurs between these two situations. However, the construction of efficient estimators is much simpler in case the restricted information set $\mathcal{H}_{i-1}$ is not too large. Therefore, from a practical point of view, alternative specifications of the restricted information set $\mathcal{H}_{i-1}$, like the one we use in Section 4, are relevant.

3 Estimation in semiparametric ACD models

3.1 Pseudo-likelihood procedures

The most basic ACD models assume that innovations are i.i.d. and exponentially distributed. This assumption has been easily rejected in several studies (see, e.g., Engle and Russell (1998) or Engle (2000)), but it can be used in a pseudo-likelihood procedure. The score-function in the ACD model with i.i.d. exponential innovations is given by

$$(\varepsilon_i - 1)\frac{\mathrm{d}}{\mathrm{d}\theta}\log(\psi_{i-1}). \tag{3.1}$$

In view of (2.1), the above score clearly satisfies the score-condition (1.1). Consequently, the pseudo-likelihood estimator based on i.i.d. exponential innovations yields consistent estimators under standard regularity conditions. However, this estimator is only efficient under fairly restrictive conditions that are discussed at the end of the current section.

One might consider enlarging the distributional class of the innovations in order to accommodate the misspecification in the exponential density. However, such an enlargement may have undesirable consequences, as we will see shortly. Two classes are widely used in the literature: the Gamma and the log-normal distributions. In both specifications one added parameter makes the exponential distribution more flexible.

Let f_λ denote the density of a normalized Gamma distribution (denoted $\Gamma(\lambda,\lambda)$), i.e.,

$$f_\lambda(x) \propto x^{\lambda-1}\exp(-\lambda x),$$

then we have

$$1 + \varepsilon_i\frac{f_\lambda'(\varepsilon_i)}{f_\lambda(\varepsilon_i)} = \lambda(1-\varepsilon_i).$$

Thus, a pseudo-likelihood procedure based on this Gamma distribution yields a score function that is proportional to (3.1). Therefore, the estimator obtained is identical to the one obtained from an exponential pseudo-likelihood procedure. The "extension" to Gamma distributions is thus void as far as a pseudo-liklihood procedure is concerned. Of course, in a parametric setting, a Gamma distribution provides a more flexible way to fit the residuals than the exponential distribution.

A second popular class of distributions is the log-normal class. The density of the normalized log-normal distribution (denoted $LN(-\frac{1}{2}\sigma^2, \sigma^2)$) is given by

$$f_{\sigma^2}(x) \propto x^{-1} \exp(-\frac{1}{2}(\log(x) + \frac{1}{2}\sigma^2)^2/\sigma^2).$$

In this class, the score function is given by

$$-\frac{1}{2} - \log(x)/\sigma^2. \tag{3.2}$$

However, the score-function (3.2) does not satisfy the score-condition (1.1) in the full semiparametric model as defined by (2.1) and (2.4). Therefore, pseudo-likelihood estimators in the ACD model based on log-normal distributions will be inconsistent. Similarly, pseudo-likelihood procedures based on other parametric classes of distributions (like the Weibull distributions) will generally yield inconsistent estimates. For the Weibull distributions this result may seem counterintuitive, since the exponential distribution is in the Weibull class. However, the inconsistency of the Weibull-based PMLE follows from the fact that the score-condition (1.1) does not hold in that case for the full semiparametric model.

Summarizing, the exponential distribution is essentially the only pseudo-distribution for which the PMLE provides consistent estimates of the ACD parameters in semiparametric settings. However, this exponential PMLE is only semiparametrically efficient under very restrictive assumptions. Indeed, the exponential PMLE is semiparametrically efficient if and only if (3.1) is proportional to the efficient score (2.10). Since, in order to achieve general efficiency, this has to hold at all f_{i-1}, we find that the exponential PMLE is efficient if and only if $(\mathrm{d}/\mathrm{d}\theta)\log\psi_{i-1}$ belongs to $\mathcal{H}_{i-1}$ and $var\{\varepsilon_i|\mathcal{H}_{i-1}\}$ is degenerate. Relaxing the pseudo-distributional assumptions in a PMLE setting may spoil the consistency of the exponential pseudo-likelihood procedure. This holds even if the relaxation includes the exponential as a special case. While there are many other examples of this effect in the literature, it is often overlooked. These considerations confirm the adaptiveness results below (2.11).

3.2 Construction of efficient semiparametric estimators

As we have seen, the often-used PMLE does not produce efficient estimators in the semiparametric ACD model. If one does not use an exponential pseudo-density, the PMLE may not even be consistent. In order to obtain semiparametrically efficient estimators, we follow standard lines that we briefly outline here. The interested reader is referred to Bickel et al. (1993), Theorem 7.8.1 and Proposition 7.8.1, or Drost, Klaassen, and Werker (1997), Theorem 3.1.

The idea is to improve an arbitrary given $\sqrt{n}$-consistent estimator towards an efficient estimator. Let $\tilde{\theta}_n$ denote this arbitrary $\sqrt{n}$-consistent estimator, for example the exponential PMLE of Section 3.1. In principle, the estimator needs also to be discretized in the sense that, for each $M > 0$ the number of elements in $\{\theta : \theta = \tilde{\theta}_n(\omega)$ for some $\omega \in \Omega$ and $\sqrt{n}|\tilde{\theta}_n - \theta_0| < M\}$ remains bounded as $n \to \infty$, where θ_0 denotes the true value of the parameter. While theoretically this discretization simplifies proofs, it is usually not required in applications. In a parametric context, where the functional form of f_{i-1} is known, an efficient estimator is obtained from a one-step Newton-Raphson improvement

$$\hat{\theta}_n = \tilde{\theta}_n + \left(\frac{1}{n} \sum_{i=1}^n \dot{l}_i(\tilde{\theta}_n) \dot{l}_i(\tilde{\theta}_n)^T \right)^{-1} \frac{1}{n} \sum_{i=1}^n \dot{l}_i(\tilde{\theta}_n). \tag{3.3}$$

Indeed, the estimator $\hat{\theta}_n$ is easily seen to have influence function

$$\left(E\dot{l}_i(\theta_0)\dot{l}_i(\theta_0)^T \right)^{-1} \dot{l}_i(\theta_0),$$

since, using (1.2) and (1.3),

$$\begin{aligned} \sqrt{n}\left(\hat{\theta}_n - \theta_0\right) &= \sqrt{n}\left(\tilde{\theta}_n - \theta_0\right) + \left(E\dot{l}_i(\theta_0)\dot{l}_i(\theta_0)^T \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{l}_i(\tilde{\theta}_n) + o_p(1) \\ &= \left(E\dot{l}_i(\theta_0)\dot{l}_i(\theta_0)^T \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{l}_i(\theta_0) + o_p(1). \end{aligned}$$

In the semiparametric model, a similar procedure is followed. The parametric score function $\dot{l}_i$ in (3.3) has to be replaced by the semiparametrically efficient score function $\dot{l}_i^*$, (2.10). Here the unknown (conditional) densities and expectations need to be consistently estimated by nonparametric methods. The exact estimation procedure is irrelevant, as long as the estimators are consistent in integrated mean-square sense. In Sections 4 and 5, we use kernel density estimators and Nadaraya-Watson regression estimators. The density of the residuals is, generally speaking, Gamma shaped. Therefore, we decided to use local bandwidth choices in the kernel estimators. To be precise about the procedure, we choose the local bandwidth such that the classical bandwidth $n^{-1/5}$ is obtained under the uniform distribution. Therefore, for each innovation, we take of the order of $n^{4/5}$ nearest neighbors into account to estimate the density. This nearest neighbor rule guarantees that the bandwidth will be smaller in regions where the density is larger. For bivariate density estimation, the rate $n^{-1/5}$ is replaced by $n^{-1/6}$.

The idea of a one-step improvement using an estimated efficient score function is rather old. Intuitively, the estimator $\tilde{\theta}_n$ brings you in a $\sqrt{n}$-neighborhood of the true value θ_0. Then, in order to obtain a locally and asymptotically efficient estimator, we need to construct an estimator with influence function

$$\left(E\dot{l}_i^*(\theta_0)\dot{l}_i^*(\theta_0)^T\right)^{-1}\dot{l}_i^*(\theta_0).$$

The local Gaussian behavior of the model (following from the LAN property), implies that the log-likelihood is approximately quadratic. The estimator $\hat{\theta}_n$ is then the maximum likelihood estimator obtained from maximizing the approximate quadratic log-likelihood following from the initial estimator $\tilde{\theta}_n$.

4 Paris Bourse: Alcatel

We illustrate the applicability of the proposed semiparametric techniques using durations observed at the Paris Bourse for transactions in Alcatel. The observations cover July and August 1996. During this period all transactions are observed. The trading system at the opening of the Paris Bourse differs from that during the rest of the day. Therefore, we delete trades within 15 minutes of the opening (compare Gourieroux and Jasiak, 1999). Simultaneous trades are aggregated: so there are no zero durations in our dataset. These simultaneous trades are usually due to large orders on one side of the market that are matched against several orders on the other side.

The mean duration in our sample is 53.2 seconds with a standard deviation of 84.8 seconds. For each time between 10:15am and 17:00pm, Figure 4.1 plots the cumulative number of trades over all days. Hence, the slope of the line reflects the average trading intensity (over all days) at a certain moment during the day. From Figure 4.1 it is clear that the average trading intensity is almost constant during the day, with lunchtime as an important exception. During lunchtime there is a clear flattening of the average trading intensity. The lower market activity is pronouncedly present in our data set and, therefore, we have to consider a mean duration function that is slightly more complicated than (2.1). We use the following specification

$$\psi_i = \alpha + \delta d_i + \beta x_i + \gamma \psi_{i-1}, \tag{4.1}$$

where d_i is an indicator for lunchtime. This extension seems to be sufficient, for the case at hand, since the trading intensity is almost constant before noon and after 2:30pm. We set $d_i = 1$ for transactions that occur between noon and 1:15 pm. Note that the exponential smoothing parameter γ will take care of a smooth transition of the "normal" intensity to the lower lunchtime intensity. By the same effect, the intensity will increase again after 1:15pm. This gradual change is seen in Figure 4.1 as the S-shaped form of the cumulative intensity around lunchtime. Engle (2000), considering IBM data, adopts a nonparametric specification of the constant in the conditional mean duration equation. There, the expected durations fluctuate in a

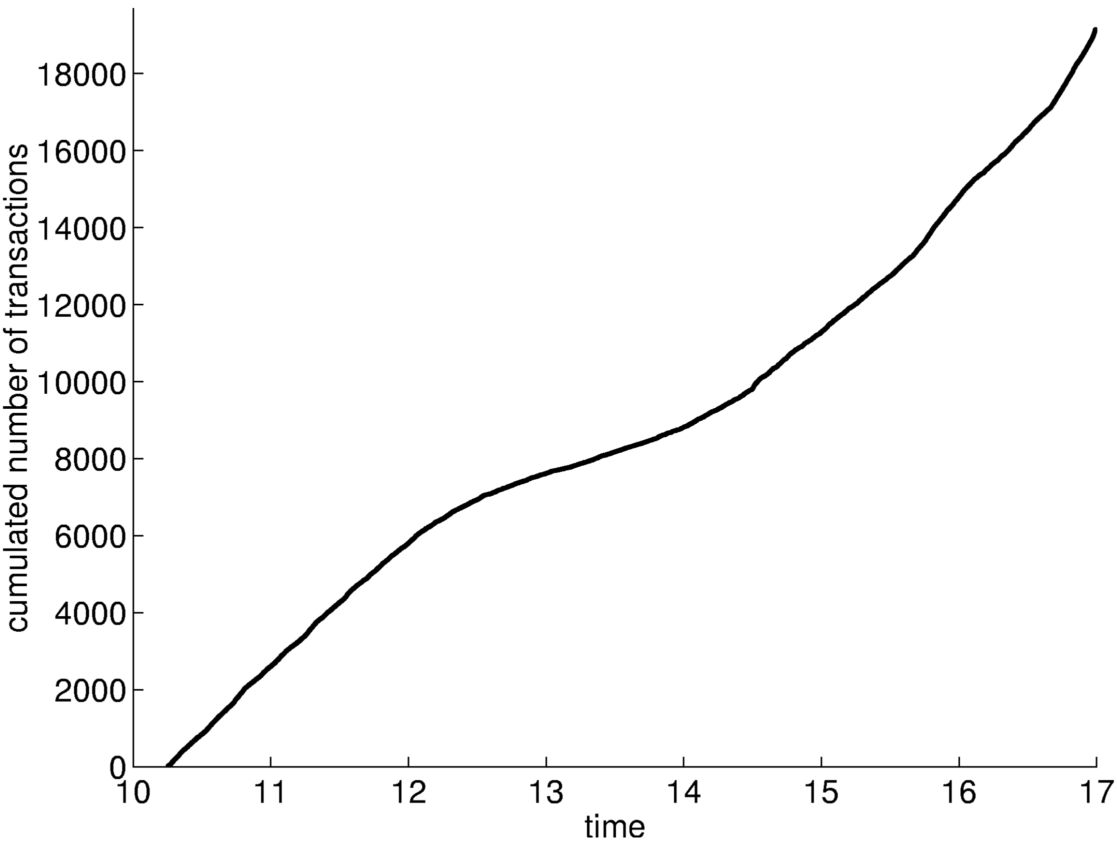

Figure 4.1: Cumulative number of trades (by daytime) for the period July 22-26, 1996.

	β		γ		α		δ	
PMLE	0.812	(0.022)	0.112	(0.016)	3.59	(0.47)	6.14	(1.05)
Martingale	0.827	(0.019)	0.107	(0.016)	3.18	(0.37)	6.01	(0.88)
Markov	0.810	(0.020)	0.155	(0.012)	2.33	(0.54)	4.55	(0.90)
IID	0.782	(0.022)	0.153	(0.013)	3.22	(0.45)	5.39	(1.10)

Table 1: Estimates of the parameters in the ACD model (2.3) for the Alcatel data based on the four procedures described in the main text. "PMLE" refers to the pseudo-likelihood method based on an exponential innovation distribution. "Martingale" is the estimator where conditional distributions are assumed to be based on past innovations. "Markov" is the efficient semiparametric estimator in case the innovation's distribution is only affected by the last innovation. "IID" refers to the optimal semiparametric estimator in the model where the innovations are assumed to be independent over transactions. Robust standard errors are reported in parentheses, see main text for details.

more pronounced way over the day and the simple approach (4.1) would fail. As long as one is interested in the parameters β and γ, this nonparametric approach could also be adopted in our current setup. It is important to note that, besides the introduction of the lunchtime dummy d_i, we do not enhance the specification of the conditional mean duration ψ_i or apply any pre-analysis transformation to the data.

We estimated the ACD model using the PMLE method and three semiparametric methods. The first estimator (indicated by "Martingale"), uses the score (2.12) where the conditional variance of the innovations is estimated by a Nadaraya-Watson nonparametric regression of the squared innovations on the previous innovation ε_{i-1} only, using the procedure outlined in Section 3.2. Such an approach is often followed in practice, even if, theoretically, the conditional variance in the optimal GMM score (2.12) depends on the whole past, i.e., $\varepsilon_{i-1}, \varepsilon_{i-2}, \ldots$. The second semiparametric estimator is based on a Markov assumption for the innovations ε_i, see Example 2.2. Again, unknown conditional densities in the efficient score function, in this case (2.10), are estimated using kernel techniques and this estimation does not affect the asymptotic semiparametric efficiency of the estimator. The estimator thus obtained is denoted "Markov". The third semiparametric estimator imposes independence of the innovations, i.e. $\mathcal{H}_i = \{\emptyset, \Omega\}$, without specifying the exact distribution (see Example 2.1). This estimator is denoted by "IID" and its theoretical properties in the general non-i.i.d. semiparametric model are unknown, but there is no reason to expect that even an elementary property as consistency is preserved. Since, the analysis of the residuals later in the present section clearly shows that the innovations are unlikely to be independent, the "IID" estimator is only given for comparison and not discussed further. Results of all four estimators for the Alcatel data are presented in Table 1.

Table 1 shows that the semiparametric procedures "Martingale" and "Markov" provide smaller standard errors than the pseudo-likelihood estimator. Generally speaking, the gain is equivalent to an increase in the number of observations by about 30%.[4] The GMM-type estimator "Martingale" and the efficient semiparametric estimator in the Markov model "Markov" behave similarly for the data at hand. A concern is a possible bias in the "Markov" estimates for the long-term levels of the durations as measured by α and δ in (4.1). Note, however, that this is in line with the fact that the constant terms in an ARCH-type stochastic volatility model cannot be estimated adaptively (see, e.g., Linton, 1993) and could be resolved along standard lines at the cost of intensive simulations.

It is known that estimates for the Fisher information in semiparametric models often have weak convergence properties. Therefore, we do not base the standard errors in Table 1 on the estimated Fisher information directly, but we apply a re-sampling technique. For each day separately, estimates of the parameters are obtained. The estimates and standard errors presented are based on the location and dispersion of the daily estimates. Assuming that the model innovations are independent over different days, this gives consistent estimates for the standard errors. We use the median and the median absolute deviation as measure for location and dispersion, respectively, to prevent a dominating effect of outlying daily estimates. The median absolute deviation is standardized such that, in case of normality, the standard deviation is obtained. Note that, on average, the daily estimates are based on approximately 500 observations. Of course, an alternative would be to use a bootstrap-type procedure, but the theoretical properties of such an approach would be difficult to establish in our non-i.i.d. situation and the computational effort involved would be enormous.

In order to assess the source of the gain of the semiparametric procedures over the pseudo-likelihood procedure, we study the residuals ε_i. In Figure 4.2, an estimate of the unconditional density of the innovations is plotted. Engle (2000) finds a similar graph (see his Figure 1) and the data suggest a non-exponential marginal distribution. To study the dependencies between the innovations, Figure 4.3 shows the autocorrelation function of the residuals, the squared centered residuals, and the log-residuals. We clearly see that both residuals and their squares are (almost) uncorrelated, while the log-residuals show a small but significantly positive autocorrelation. The proposed semiparametric procedure takes such dependencies effectively into account. The autocorrelation of log-residuals is only about 0.094. Apparently, such low dependencies still show up in the efficiency gains for the semiparametric estimators. As mentioned before, in our situation the gain is of the order of 30% of the observations.

[4] This number is obtained as the average relative efficiency of the "Martingale" estimator and the "Markov" estimator with respect to the PMLE.

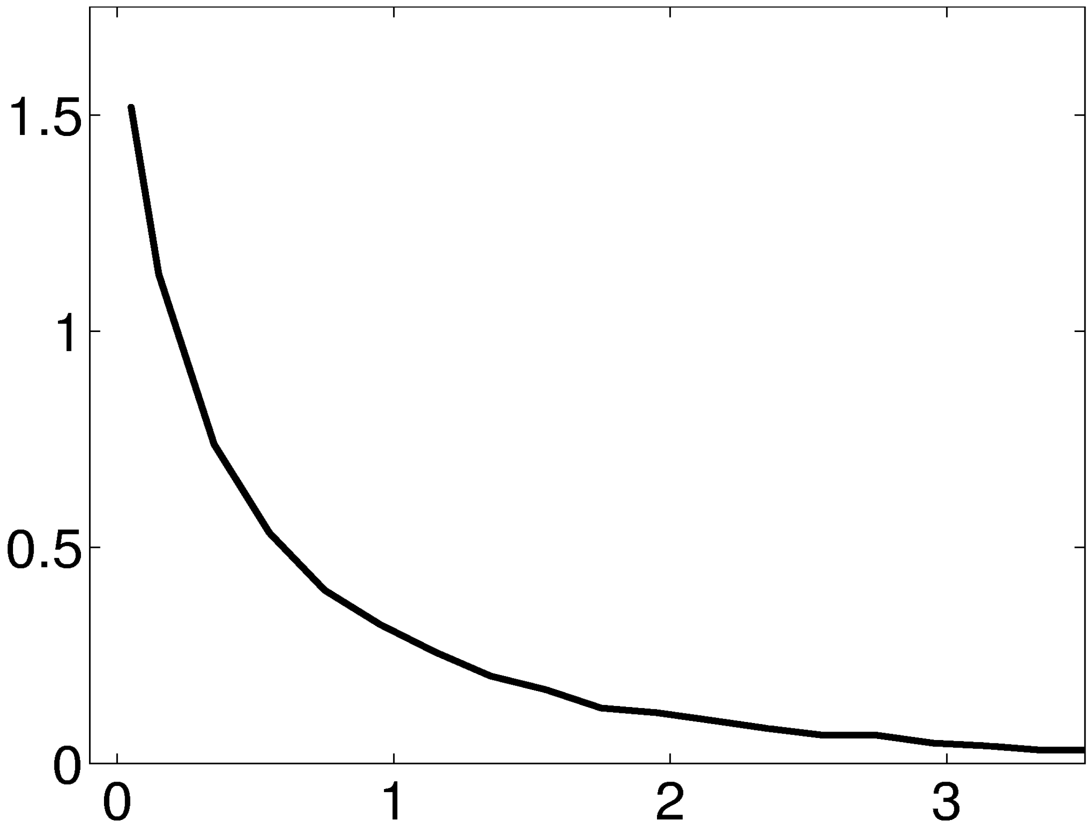

Figure 4.2: Estimate of unconditional density of innovations.

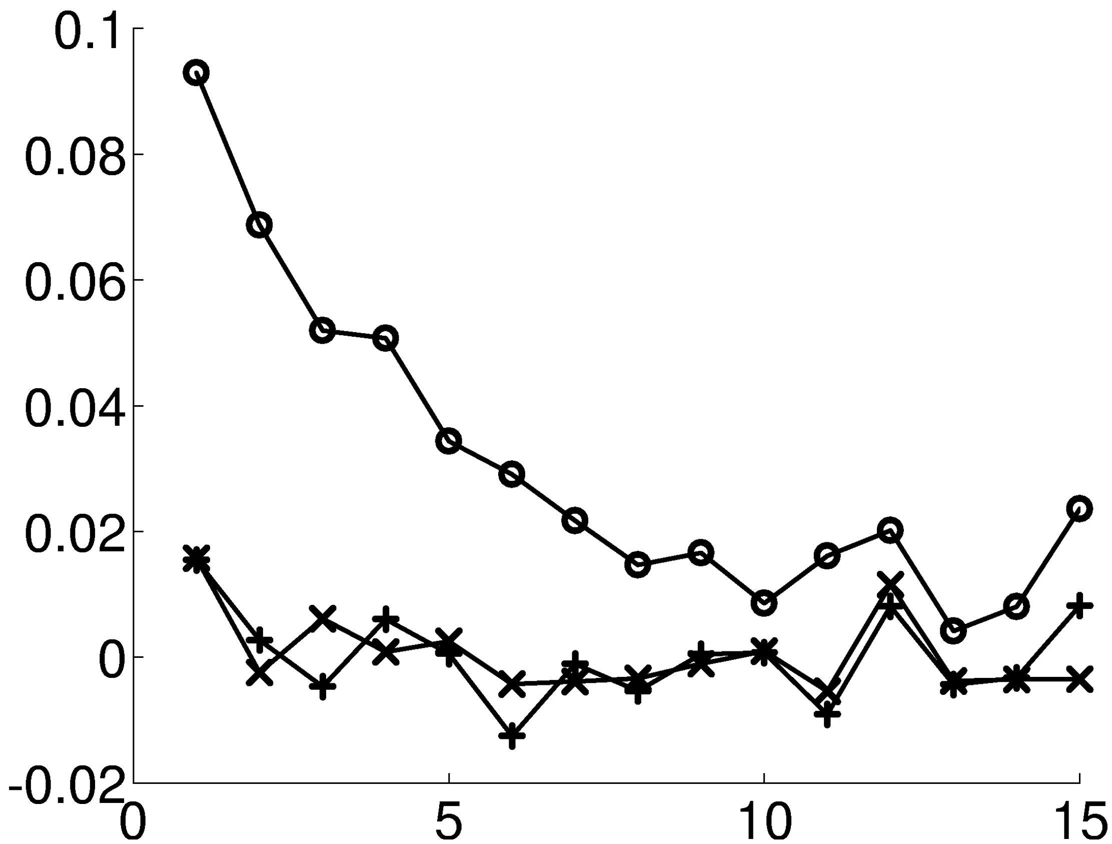

Figure 4.3: ACF of residuals ($\bigcirc - \bigcirc - \bigcirc$), squared centered residuals ($\times - \times - \times$), and log-residuals ($+ - + - +$).

5 Simulation results

The results in the previous section indicate that even small dependencies in the innovations as measured, e.g., by the autocorrelation of the log-residuals can induce sizeable efficiency gains of efficient semiparametric procedures over pseudo-likelihood procedures. In the current section we investigate this in more detail. Thereto, we consider parametric models that mimic the most salient features of the Alcatel data. We do not advocate to use these parametric models as an alternative to the semiparametric models introduced in Section 2.2, since misspecification is quite likely and may adversely affect the estimators. The parametric models in this section are merely used to confirm the properties of the semiparametric estimators in realistic settings.

The residuals of the Alcatel durations in Section 4 show some delicate dependencies. Clearly, the model specification requires that the residuals are uncorrelated. Squared residuals also appear to be uncorrelated, while logarithmic residuals show some weak but significant first-order autocorrelation. An extension of the classical Gamma (including the exponential) or log-normal specifications incorporating these stylized features, is obtained by making the parameters of those distributions time-varying. As an example, consider the following possible specifications

$$\varepsilon_i \sim \Gamma(\sigma_{i-1}^{-2}, \sigma_{i-1}^{-2}), \tag{5.1}$$

or

$$\varepsilon_i \sim LN(-\frac{1}{2}\log(1+\sigma_{i-1}^2), \log(1+\sigma_{i-1}^2)), \tag{5.2}$$

with

$$\sigma_{i-1}^2 = 0.1 + 0.9\varepsilon_{i-1}. \tag{5.3}$$

Note that for both specifications the conditional variance of the innovations is indeed given by σ_{i-1}^2. Clearly, the above specifications are not the only parametric ones that generate dependence structures comparable to those found in the Alcatel data. Therefore, we advocate the use of a semiparametric technique for econometric analysis of the structural parameters in the specification of the conditional expected duration ψ_i. This seems all the more reasonable since a parametrically misspecified model of the innovation's distribution does not produce consistent pseudo-likelihood estimates in general. As has been pointed out before, this holds even if the parametric specification includes the i.i.d. exponential specification for which pseudo-likelihood procedures are consistent.

We present results for the same four estimators as used in the analysis of the Alcatel data. The first estimator ("PMLE") is the exponential PMLE. For the second estimator ("Martingale"), the conditional variance of the innovations may depend in an arbitrary way on the past. The third estimator ("Markov") is based on the efficient semiparametric score (2.10) and assumes that the innovations follow a Markov process with unknown transition density. The final estimator ("IID") is the efficient semiparametric estimator in case the innovations are i.i.d., see Example 2.1. The

IID Exponential Innovations						
	Exact scores					
	β		γ		α	
PMLE	0.788	(0.0121)	0.099	(0.0055)	4.86	(0.368)
Martingale	0.788	(0.0121)	0.099	(0.0055)	4.86	(0.368)
Markov	0.788	(0.0121)	0.099	(0.0055)	4.86	(0.368)
IID	0.788	(0.0121)	0.099	(0.0055)	4.86	(0.368)
	Estimated scores					
	β		γ		α	
PMLE	0.788	(0.0121)	0.099	(0.0055)	4.86	(0.368)
Martingale	0.788	(0.0125)	0.098	(0.0056)	4.82	(0.386)
Markov	0.791	(0.0135)	0.096	(0.0058)	4.72	(0.424)
IID	0.790	(0.0132)	0.098	(0.0058)	4.78	(0.397)

Table 2: Simulation results for the ACD model (2.3) with i.i.d. exponential innovations. See Table 1 for an explanation of the terminology used.

true values in (2.3) are $\alpha = 4.50$, $\beta = 0.80$, and $\gamma = 0.1$ and we consider both the Gamma specification (5.1) and the log-normal specification (5.2). The daily number of observations is, in accordance with the average in the Alcatel data, fixed at 500. The computational effort in the simulations is substantial. Therefore, the number of replications is limited to $2,500$. Again, we present location and dispersion estimates that are based on robust estimates, i.e. the median and the median absolute deviation. The reported standard errors are multiplied with $\sqrt{2,500/43}$ in order to make them comparable to the empirical results of Section 4.

The simulation results for the Gamma and the log-normal specification are presented in Tables 3-4. For reference, we also present the results in case the innovations are independently and identically exponentially distributed, see Table 2. We present both the estimation results based on exact scores (i.e., as appropriate for the data generating process at hand) and results based on estimated scores. From this, we can then infer the theoretical semiparametric efficiency gain and the effect of the non-parametric density and regression function estimates. Table 2 presents the results in the ideal situation of i.i.d. exponential innovations ε_i. As discussed in Section 3, all four estimators are efficient (even adaptive) in this case. Indeed, the scores used by all estimators are the same and, consequently, when using exact scores, the estimators are identically equal to the PMLE. In case the score functions are estimated, the estimators still behave theoretically the same. The simulation results confirm this as there is little variation with respect to standard errors. One may notice a slight increase in the variation for the semiparametric "Markov" estimator which is caused by the non-parametric conditional density estimation therein.

To examine the effect of dependencies on the performance of the estimators, we

Conditional Gamma Innovations						
	Exact scores					
	β		γ		α	
PMLE	0.804	(0.0140)	0.088	(0.0073)	4.48	(0.367)
Martingale	0.800	(0.0113)	0.093	(0.0052)	4.60	(0.328)
Markov	0.800	(0.0113)	0.093	(0.0052)	4.60	(0.328)
	Estimated scores					
	β		γ		α	
PMLE	0.804	(0.0140)	0.088	(0.0073)	4.48	(0.367)
Martingale	0.807	(0.0110)	0.090	(0.0052)	4.37	(0.315)
Markov	0.813	(0.0131)	0.086	(0.0059)	4.18	(0.372)
IID	0.803	(0.0143)	0.090	(0.0068)	4.57	(0.399)

Table 3: Simulation results for the ACD model (2.3) with innovation structure (5.1) where $\sigma_{i-1}^2 = 0.1 + 0.9\varepsilon_{i-1}$. See Table 1 for an explanation of the terminology used.

first consider the conditional Gamma innovations in (5.1). In this case, the PMLE and the "Martingale" and "Markov" semiparametric estimators provide consistent estimates. There is no guarantee (known to us) that the IID semiparametric estimator is consistent in this setting with dependent innovations. Of course, calculations with exact scores cannot be performed for this estimator. The results based on exact scores, show that the theoretical standard errors of the PMLE are somewhat larger than those of the two consistent semiparametric estimators. This confirms the results of Section 3 as the conditions under which the PMLE provides efficient estimates are not met in the present simulation where $var\{\varepsilon_i|\mathcal{H}_{i-1}\}$ is non-degenerate. Note, however, that since the innovations are conditionally Gamma distributed, the "Martingale" and "Markov" semiparametric estimators are theoretically equal. However, the density estimation required in the implementation of the "Markov" semiparametric estimator increases its variability to the level of the PMLE, while the "Martingale" estimator retains its theoretical variability.

The former two simulations are still quite specific since, asymptotically, the "Martingale" and "Markov" semiparametric estimators coincide. We, therefore, also conducted the analysis using conditionally log-normally distributed innovations as in (5.2). As noted before, the log-normal distribution is not suited as a pseudo-distribution in a PMLE procedure, since such an estimator would generally be inconsistent. However, it is informative to investigate the effect of log-normal innovations on the simulation results, see Table 4. Again, the last semiparametric estimator (based on an i.i.d.-ness assumption on the innovations) does not produce consistent estimates in this case. This is clearly visible from the table, both using exact scores and estimated scores. As before, the "Martingale" and "Markov" semiparametric estimators show efficiency gains over the exponential PMLE, however these estimators are no longer asymptot-

Conditional log-normal Innovations						
	Exact scores					
	β		γ		α	
PMLE	0.806	(0.0141)	0.088	(0.0070)	4.47	(0.379)
Martingale	0.795	(0.0137)	0.097	(0.0064)	4.65	(0.371)
Markov	0.796	(0.0112)	0.098	(0.0051)	4.66	(0.316)
	Estimated scores					
	β		γ		α	
PMLE	0.806	(0.0141)	0.088	(0.0070)	4.47	(0.379)
Martingale	0.810	(0.0121)	0.087	(0.0058)	4.26	(0.339)
Markov	0.817	(0.0112)	0.085	(0.0056)	4.06	(0.324)
IID	0.912	(0.0084)	0.035	(0.0039)	2.32	(0.242)

Table 4: Simulation results for the ACD model (2.3) with innovation structure (5.2) where $\sigma_{i-1}^2 = 0.1 + 0.9\varepsilon_{i-1}$. See Table 1 for an explanation of the terminology used.

ically equivalent. The table shows a slight improvement of the "Markov" estimator. The gains of the efficient semiparametric procedures over the standard exponential PMLE are, as for the Alcatel data, roughly in the order of magnitude of 30% of the number of observations.

Note that the standard errors for all simulations differ somewhat from those found for the Alcatel data. This suggests that in the Alcatel data even more complicated dependencies than those studied in this section play a role. Clearly, the use of semiparametric techniques avoids misspecification problems inherently present when using parametric models. Note that, the simulation results for the "Martingale" and "Markov" semiparametric estimators are quite similar in all cases. Apparently, for the specifications chosen in this section, the respective scores (2.10) and (2.12) are close.

Summarizing, the simulations confirm that significant efficiency gains may be obtained from the use of semiparametric procedures. We prefer the theoretically optimal semiparametric estimators. Even if large numbers of observations are available for the study of intra-day durations, the semiparametric procedures allow for much more precise empirical analysis and prediction. Moreover, with large datasets the distortions induces by the nonparametric density estimation are likely to disappear. Recall that our results are based on only a moderate sample of size 500.

6 Concluding remarks

We discussed optimal estimation in semiparametric duration models. The models differ in the specification of the possible dependencies between the innovations. These specifications range from the case where innovations are i.i.d. with unknown density to completely arbitrary dependencies that only impose an identifying martingale restriction. For these specifications, we derived the efficient score functions for the parameters of interest that govern the conditional expected duration. We also showed that the often used exponential PMLE is only efficient under very restrictive conditions and that the other PMLE's (e.g., based on the log-normal or Weibull distribution) are not consistent. We showed that an easily implementable semiparametric estimator allows for significant (comparable to 30% of the observations) efficiency gains. In order to find a possible explanation for this phenomenon, we set up a simulation experiment with time-varying parameters in the innovation's distribution. The stylized features of the Alcatel data for our observation period are mimicked in this experiment. These simulations confirm the fact that the semiparametric procedures outperform pseudo-likelihood procedures.

References

Bickel, P.J., C.A.J. Klaassen, Y. Ritov, and J.A. Wellner (1993), *Efficient and Adaptive Statistical Inference for Semiparametric Models*, John Hopkins University Press, Baltimore.

Drost, F.C., and C.A.J. Klaassen (1997), Efficient estimation in semiparametric GARCH models, *Journal of Econometrics*, **81**, 193–221 .

Drost, F.C., C.A.J. Klaassen, and B.J.M. Werker (1997), Adaptive estimation in time-series models, *The Annals of Statistics*, **25**, 786–818.

Drost, F.C., C.A.J. Klaassen, and B.J.M. Werker (2000), Semiparametric lower bounds in time series models with non-I.I.D. error structures, work in progress.

Engle, R.F., and J.R. Russell (1998), Autoregressive conditional duration: A new model for irregularly spaced transaction data, *Econometrica*, **66**, 1127–1162.

Engle, R.F. (2000), The econometrics of ultra-high-frequency data, *Econometrica*, **68**, 1–22.

González-Rivera, G. (1997), A Note on Adaptation in GARCH Models, *Econometric Reviews*, **16**, 55–68.

Gourieroux, C., and J. Jasiak (1999), Nonlinear Autocorrelograms; an Application to Intertrade Durations, working paper.

Gourieroux, C., E. Ghysels, and J. Jasiak (1999), Stochastic volatility duration models, preprint.

Koul, H.L., and A. Schick (1997), Efficient estimation in nonlinear autoregressive time series models, *Bernoulli*, **3**, 247-277.

Kreiss, J.-P. (1987a), On adaptive estimation in stationary ARMA processes, *The Annals of Statistics*, **15**, 112–133.

Kreiss, J.-P. (1987b), On adaptive estimation in autoregressive models when there are nuisance functions, *Statistics & Decisions*, **5**, 59-76.

Linton, O. (1993), Adaptive estimation in ARCH models, *Econometric Theory*, **9**, 539–569.

Newey, W.K. (1990), Semiparametric efficiency bounds, *Journal of Applied Econometrics*, **5**, 99–135.

Steigerwald, D.G. (1992), Adaptive estimation in time series regression models, *Journal of Econometrics*, **54**, 251–275.

Stein, C. (1956), Efficient nonparametric testing and estimation, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 187–195.

Van der Vaart, A. (1998), Asymptotic statistics, Cambridge University Press, Cambridge.

Wefelmeyer, W. (1996), Quasi-likelihood models and optimal inference, *The Annals of Statistics*, **24**, 405–422.